DOCUMENT RESUME

ED 264 261 TM 850 745

AUTHOR Samejima, Fumiko
TITLE Development and Application of Methods for Estimating Operating Characteristics of Discrete Test Item Responses without Assuming any Mathematical Form.
PUB DATE Jul 82
NOTE 39p.; In: Item Response Theory and Computerized Adaptive Testing Conference Proceedings (Wayzata, MN, July 27-30, 1982) (TM 850 744).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Computer Oriented Programs; *Estimation (Mathematics); Item Analysis; *Latent Trait Theory; *Psychometrics; *Research Methodology; Scores; Statistical Analysis; Statistical Studies; *Test Items
IDENTIFIERS *Unidimensional Scaling

ABSTRACT

In latent trait theory the latent space, or space of the hypothetical construct, is usually represented by some unidimensional or multi-dimensional continuum of real numbers. Like the latent space, the item response can either be treated as a discrete variable or as a continuous variable. Latent trait theory relates the item response to the latent trait in terms of the operating density characteristic when the item response is continuous and by means of the operating characteristic when the item response is discrete. This paper is concerned solely with the unidimensional latent space and the discrete response level. An outline for a series of methods and approaches is introduced for estimating the operating characteristics of discrete item responses without assuming any mathematical form. Observations concerning computer programs and model validation, word/phrase comprehension tests and scale construction, and information loss caused by asumptotes are discussed. Numerous tables and figures are included, as well as three pages of references. (PN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

Development and Application of Methods for Estimating Operating Characteristics of Discrete Test Item Responses without Assuming Any Mathematical Form

Fumiko Samejima
University of Tennessee

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☒ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ED264261

In latent trait theory the latent space, or space of the hypothetical construct, is usually represented by some unidimensional or multidimensional continuum of real numbers. Many researchers have concentrated their effort on unidimensional spaces, although there have been some models developed in the multidimensional latent space (Samejima, 1974a, 1974b), and this will be an important future orientation of research. Like the latent space, the item response can either be treated as a discrete variable or as a continuous variable. Although most researchers have solely treated discrete responses, there are certain papers which have dealt with continuous responses (Samejima, 1973a, 1974a).

Latent trait theory relates the item response to the latent trait in terms of the operating density characteristic when the item response is continuous, and by means of the operating characteristic when the item response is discrete. By the operating density characteristic of the continuous item response is meant the conditional density function of a particular item response, given the latent trait. This operating density characteristic is replaced by the operating characteristic of the item response on the discrete response level, which is the conditional probability function of the item response, given the latent trait. The discrete response level includes the nominal response level, the graded response level, and the dichotomous response level (Samejima, 1972).

On the nominal response level, response categories are not explicitly ordered. Using information given by the incorrect as well as the correct answer of the multiple-choice test item provides a good example of the nominal response level. Bock (1972) proposed a multinomial response model, which is applicable to the multiple-choice test item when the effect of random guessing is negligibly small. Samejima (1979b) proposed a family of models for the multiple-choice test item, which accounts both for the information provided by the incorrect answers and for the effect of random guessing.

The graded response level is further categorized into the homogeneous and heterogeneous cases. Samejima (1969, 1972) has proposed the normal ogive model and the logistic model developed for the homogeneous case of the graded response level. In a special case of the graded response level where the item response is binary, e.g., correct or incorrect in ability or achievement measurement,

TM 850 745

Table 1

Notation

θ	: unidimensional latent trait
g	: item of Old Test (=1, 2, ..., n)
k_g	: discrete item response to item g
$P_{k_g}(\theta)$	: operating characteristic of k_g
$A_{k_g}(\theta)$	: basic function of k_g
V	: response pattern or vector of n discrete item responses
$L_V(\theta)$	: likelihood function of θ given response pattern
$P_V(\theta)$	: operating characteristic response pattern V
$I_{k_g}(\theta)$	: item response information function of discrete response k_g
$I_g(\theta)$	: item information function of item g
$I_V(\theta)$	: response pattern information function of V
$I(\theta)$	: test information function of Old Test
$\hat{\theta}_V$, $\hat{\theta}$	: maximum likelihood estimate of θ
ε	: error of maximum likelihood estimate $\hat{\theta}$
x_g	: graded item response to item g (= 0, 1, ..., m) (special case of k_g)
$P_{x_g}(\theta)$	: operating characteristic of x_g
a_g	: discrimination parameter of item g
b_{x_g}	: difficulty parameter of x_g
$P_g(\theta)$	: item characteristic function of item g
b_g	: difficulty parameter of item g when g is binary
$\underline{\theta}$, $\overline{\theta}$	: lower and upper endpoints of arbitrarily chosen subinterval of θ
$\underline{\underline{\theta}}$, $\overline{\overline{\theta}}$	: lower and upper endpoints of the interval of θ for which $P_g(\theta)$ of Type A is strictly increasing in θ
$\sigma(\theta)$	: standard error of maximum likelihood estimate $\hat{\theta}_V$ given θ
$f(\theta)$	: density function of latent trait θ
τ	: transformed latent trait, which is strictly increasing in θ
$P^*_{k_g}(\tau)$	: operating characteristic of k_g defined on τ
$P^*_V(\tau)$	: operating characteristic of V defined on τ

- continued on next page -

Table 1 (continued)

Notation

$I^*_{k_g}(\tau)$	: item response information function defined on τ
$I^*_g(\tau)$	: item information function defined on τ
$I^*_V(\tau)$	: response pattern information function defined on τ
$I^*(\tau)$	: test information function defined on τ
$\hat{\tau}_V$, $\hat{\tau}$	: maximum likelihood estimate of τ based on V
$\underline{\tau}$, $\overline{\tau}$	: lower and upper endpoints of the interval corresponding to $(\underline{\theta},\overline{\theta})$
τ_0	: value of τ corresponding to the origin of θ
$f^*(\tau)$	: density function of transformed latent trait τ
C	: constant square root of test information defined on τ
α_k	: (k = 0, 1, ..., m) coefficients of polynomial approximating the square root of the test information function of θ
α^*_k	: (k = 0, 1, ..., m+1) coefficients of polynomial transforming θ to τ
ε^*	: error of maximum likelihood estimate $\hat{\tau}$
$\dot{\varepsilon}(\hat{\tau})$	: sample linear regression of ε^* on $\hat{\tau}$
$g(\hat{\tau})$	: density function of $\hat{\tau}$
s	: examinee (= 1, 2, ..., N)
$\hat{\tau}_s$	: maximum likelihood estimate of $\hat{\tau}$ assigned to examinee $\underline{s}$
h	: test item whose operating characteristics are to be discovered, or "unknown test item"
k_h	: discrete item response to unknown test item $\underline{h}$
$P_{k_h}(\theta)$	: operating characteristic of k_h
$f_{k_h}(\theta)$	: density function of θ for the subgroup of examinees who share the discrete item responses k_h to item $\underline{h}$
N_{k_h}	: number of examinees who share the discrete item response k_h to item $\underline{h}$ in the sample
$g_{k_h}(\hat{\tau})$	: density function of $\hat{\tau}$ for the subgroup of examinees who share the discrete item response k_h to item $\underline{h}$
$\phi_{k_h}(\tau\|\hat{\tau})$	: conditional density function of τ, given $\hat{\tau}$
$\xi_{k_h}(\tau,\hat{\tau})$	: joint density function of τ and $\hat{\tau}$
$w(\hat{\tau}_s)$	: weight assigned to $\hat{\tau}_s$
$p(s\varepsilon k_h)$	: probability with which the examinee $\underline{s}$ belongs to the item response subgroup k_h

Table 2
Fundamental Formulae

$$V = (k_1, k_2, \ldots, k_g, \ldots, k_n)' \qquad [1]$$

$$P_V(\theta) = \prod_{k_g \varepsilon V} P_{k_g}(\theta) \qquad \text{(by local independence)} \qquad [2]$$

$$\frac{\partial}{\partial\theta} \log L_V(\theta) = \frac{\partial}{\partial\theta} \log P_V(\theta) = \sum_{k_g \varepsilon V} A_{k_g}(\theta) = 0$$

(likelihood equation) [3]

$$A_{k_g}(\theta) = \frac{\partial}{\partial\theta} \log P_{k_g}(\theta) \qquad [4]$$

$$I_{k_g}(\theta) = -\frac{\partial}{\partial\theta} A_{k_g}(\theta) \qquad [5]$$

$$I_g(\theta) = E[I_{k_g}(\theta)|\theta] = \sum_{k_g} I_{k_g}(\theta)\, P_{k_g}(\theta)$$

$$= \sum_{k_g} \left[\frac{\partial^2}{\partial\theta^2} P_{k_g}(\theta)\right]^2 [P_{k_g}(\theta)]^{-1} \qquad [6]$$

$$I_V(\theta) = -\frac{\partial^2}{\partial\theta^2} \log P_V(\theta) = \sum_{k_g \varepsilon V} I_{k_g}(\theta) \qquad [7]$$

$$I(\theta) = E[I_V(\theta)|\theta] = \sum_V I_V(\theta)\, P_V(\theta) \qquad [8]$$

$$I(\theta) = \sum_{g=1}^{n} I_g(\theta) \qquad \text{(derived from Equations 6, 7, and 8)} \qquad [9]$$

$$\tau = \tau(\theta) \qquad [10]$$

$$P^*_{k_g}[\tau(\theta)] = P_{k_g}(\theta) \qquad [11]$$

$$P^*_V[\tau(\theta)] = P_V(\theta) \qquad [12]$$

- continued on the next page -

Table 2 (continued)
Fundamental Formulae

$$I^*_{k_g}(\tau) = I_{k_g}(\theta)\left[\frac{d\theta}{d\tau}\right]^2 - \frac{\partial}{\partial\theta}\log P_{k_g}(\theta)\cdot\frac{d^2\theta}{d\tau^2} \qquad [13]$$

$$I^*_g(\tau) = I_g(\theta)\left[\frac{d\theta}{d\tau}\right]^2 \qquad [14]$$

$$I^*_V(\tau) = I_V(\theta)\left[\frac{d\theta}{d\tau}\right]^2 - \frac{\partial}{\partial\theta}\log P_V(\theta)\cdot\frac{d^2\theta}{d\tau^2} \qquad [15]$$

$$I^*(\tau) = I(\theta)\left[\frac{d\theta}{d\tau}\right]^2 \qquad [16]$$

$$\begin{cases} \underline{\tau} = \tau(\underline{\theta}) \\ \bar{\tau} = \tau(\bar{\theta}) \end{cases} \qquad [17]$$

$$\int_{\underline{\tau}}^{\bar{\tau}} [I^*_g(\tau)]^{1/2}\, d\tau = \int_{\underline{\theta}}^{\bar{\theta}} [I_g(\theta)]^{1/2}\, d\theta$$

(constancy of square root of item information) [18]

$$\int_{\underline{\tau}}^{\bar{\tau}} [I^*(\tau)]^{1/2}\, d\tau = \int_{\underline{\theta}}^{\bar{\theta}} [I(\theta)]^{1/2}\, d\theta$$

(constancy of square root of test information) [19]

positive or negative in attitude measurement, and so forth, the operating characteristic of the positive response is called the item characteristic function (Lord & Novick, 1968, chap. 16). On this dichotomous response level, mathematical models such as the Rasch (1960) model, the normal ogive model (Lord, 1952; Tucker, 1951), the logistic model and its 3-parameter version (Birnbaum, 1968) have been proposed and used.

The present paper is concerned solely with the unidimensional latent space and the discrete response level. Samejima (1977c, 1977e, 1978a, 1978b, 1978c, 1978d, 1978e, 1978f, 1980a, 1980b) has developed a series of methods and approaches for estimating the operating characteristics of discrete item responses without assuming any mathematical form. The outline of those methods and approaches will be introduced in the present paper. The direct estimation of the operating characteristics of discrete item responses has also been attempted by Lord (1969, 1970), in which the estimation of the true test score distribution

of each subgroup of examinees plays an essential role. Levine (1980) has also developed a method for a similar purpose, in which he has utilized eigenfunctions effectively. Some observations concerning model validation, scale construction, and information loss will also be made.

Notation and Fundamental Formulae

Table 1 presents the set of basic symbols used in the present study. Assume that there is a set of n items, or Old Test, whose operating characteristics are known, and let g denote an arbitary item of the Old Test. An "unknown" test item i.e., any test item whose operating characteristics are to be discovered, is denoted by h, to distinguish it from item g. Also assume that the Old Test has a substantially large amount of test information at any point of the interval of θ of interest.

Table 2 presents the set of fundamental formulae which relate these mathematical symbols with one another. They are the basis of the present study.

Asymptotic Properties of the Maximum Likelihood Estimate and Transformation of Latent Trait

The well-known property of the maximum likelihood estimate--under a general condition its distribution converges to normality--is fully utilized in the present study. The proof of this theorem is given in many statistics books when the observations are taken from identical distributions (e.g., Kendall & Stuart, 1961). It can easily be expanded to the general case where the test items are not necessarily equivalent (Samejima, 1976).

Let ε be the error in the maximum likelihood estimate, $\hat{\theta}_V$, of the latent trait θ, so that

$$\hat{\theta}_V = \theta + \varepsilon \ . \qquad [20]$$

Thus, $\hat{\theta}_V$ is asymptotically and conditionally unbiased and normally distributed with $\sigma(\theta)$ as the second parameter, such that

$$E(\hat{\theta}_V|\theta) = \theta \qquad [21]$$

and

$$\sigma(\theta) = [I(\theta)]^{-1/2} \ . \qquad [22]$$

The speed of convergence of this conditional distribution to normality has been observed by monte carlo studies (Samejima, 1976, 1977a, 1977b, 1977c, 1979a). In three studies, except for the last one in which the Constant Information Model was adopted, either the normal ogive model on the dichotomous response level or the one on the graded response level was used. The item characteristic function, $P_g(\theta)$, of the first model is given by

$$P_g(\theta) = [2\pi]^{-1/2} \int_{-\infty}^{a_g(\theta - b_g)} e^{-u^2/2} \, du \ , \qquad [23]$$

where a_g (>0) and b_g are the item discrimination and difficulty parameters, respectively. For the operating characteristic of the graded response x_g (= 0, 1, ..., m_g) in the second model and

$$P_{x_g}(\theta) = [2\pi]^{-\frac{1}{2}} \int_{a_g(\theta-b_{x_g+1})}^{a_g(\theta-b_{x_g})} \qquad [24]$$

where

$$-\infty = b_0 < b_1 < \ldots < b_{m_g} < b_{m_g+1} = \infty \; . \qquad [25]$$

These different monte carlo studies have shown that the speed of convergence is high, i.e., the conditional distribution of $\hat{\theta}_V$, given θ, can be approximated by $N(\theta, I(\theta)^{-\frac{1}{2}})$ even when the number of test items is relatively small. The benefit of this finding will be enhanced if the test information function, $I(\theta)$, assumes a constant value for the interval of interest, for it will simplify the subsequent mathematics. In computerized adaptive testing the constant standard error of estimation can be obtained easily for the interval of θ of interest (Samejima, 1977b). This can be accomplished by using a set amount of test information as the criterion for terminating the presentation of new test items for each individual examinee. In so doing, weakly parallel tests (Samejima, 1977d) are produced in each session, with the same accuracy of estimation everywhere within the interval of θ that is of interest.

When this is not the case, the latent trait θ can be transformed to τ, which provides constant test information (cf. Samejima, 1980a). Let C be the desired constant value of the square root of the test information function for the interval $[\underline{\tau}, \overline{\tau}]$. Then, the transformation of θ to τ is given by

$$\tau(\theta) = C^{-1} \int_{-\infty}^{\theta} [I(u)]^{\frac{1}{2}} du + \tau_0 \; , \qquad [26]$$

where τ_0 is an arbitrary constant, or the value of τ which corresponds to the origin of θ. Note, however, that in order to approximate the conditional distribution of the maximum likelihood estimate $\hat{\tau}_V$, given the latent trait τ, it is necessary that in spite of its irregularity, the amount of test information defined for the original latent trait θ be substantially large throughout the interval of θ of interest. By virtue of the transformation-free character of the maximum likelihood estimate, the new maximum likelihood estimate, $\hat{\tau}_V$, can be obtained through the same transformation, such that

$$\hat{\tau}_V = \tau(\hat{\theta}_V) \; . \qquad [27]$$

The transformation of θ to τ, which is given by Equation 26, may be rather

complicated, for it includes the integration of the square root of the test information function $I(\theta)$, given by Equation 9 (see Table 2). The process will be largely simplified if the square root of the test information function is approximated by a polynomial obtained by the method of moments (Elderton & Johnson, 1969). It has been observed (Samejima & Livingston, 1979) that the polynomial of a specified degree, m, which is obtained by the method of moments, provides the least squares solution. By the use of the method of moments, rounding error is also reduced, for it does not include the inversion of an ill-conditioned matrix. With such a polynomial, for the square root of the test information function

$$[I(\theta)]^{1/2} \doteq \sum_{k=0}^{m} \alpha_k \theta^k \ . \qquad [28]$$

can be written. Substituting Equation 28 into Equation 27 gives

$$\tau(\theta) \doteq C^{-1} \sum_{k=0}^{m} \alpha_k (k+1)^{-1} \theta^{k+1} + \tau_0$$

$$= \sum_{k=0}^{m+1} \alpha_k^* \theta^k \ , \qquad [29]$$

where

$$\alpha_k^* \begin{cases} = \tau_0 & k = 0 \\ = (Ck)^{-1} \alpha_{k-1} & k = 1,2, \ldots, m+1 \ . \end{cases} \qquad [30]$$

Approach to the Joint Density Function of τ and $\hat{\tau}_V$

When the operating characteristics of the test items of the Old Test satisfy the unique maximum condition (Samejima, 1969, 1972), the maximum likelihood estimate of the latent trait θ can be obtained for each of those examinees to whom the Old Test was administered. Perhaps the most naive way of obtaining the estimated operating characteristics of a discrete item response is to take the simple ratio of the frequency of the examinees who share the response to the total frequency, by setting small subintervals of $\hat{\theta}_V$. This method has certain serious theoretical problems, however, in addition to the fact that it requires a large sample size and, in general, a substantial amount of computer time in obtaining $\hat{\theta}_V$. Generally speaking, the regression of θ on $\hat{\theta}_V$ is not even linear. Thus, even if the maximum likelihood estimate $\hat{\theta}_V$ is approximately conditionally unbiased, given the latent trait θ and constant test information for the interval of θ of interest, the resultant estimated operating characteristic will be stretched unevenly. For the expectation of $\hat{\theta}_V$

$$E(\hat{\theta}_V) = E(\theta) \ , \qquad [31]$$

when the maximum likelihood estimate $\hat{\theta}_V$ is approximately conditionally unbiased, given θ. The general formula for the mth moment of $\hat{\theta}_V$ is given by

$$E[\hat{\theta}_V - E(\hat{\theta}_V)]^m = \sum_{r=0}^{m} \binom{m}{r} E[\{\theta - E(\theta)\}^{m-r} E\{(\hat{\theta}_V - \theta)^r | \theta\}] \qquad [32]$$

(cf. Samejima, 1977c). From Equation 32

$$Var(\hat{\theta}_V) = Var(\theta) + E[Var(\hat{\theta}_V|\theta)] \geq Var(\theta) \ , \qquad [33]$$

is obtained for the variance of $\hat{\theta}_V$, ignoring the covariance term for the case where the conditional unbiasedness of $\hat{\theta}_V$ approximately holds. For these reasons, it is desirable to discover methods which ameliorate these deficiencies.

The group of examinees of the sample can be categorized into subgroups depending upon their item scores, k_h, to the "unknown" item h. If the density function, $\tilde{f}_{k_h}(\theta)$, of each discrete item score k_h is estimated, then the estimated operating characteristics, $\tilde{P}_{k_h}(\theta)$, can be represented by

$$\tilde{P}_{k_h}(\theta) = N_{k_h} \cdot \tilde{f}_{k_h}(\theta) \left[\sum_j N_j \cdot \tilde{f}_j(\theta)\right]^{-1} \ , \qquad [34]$$

where N_{k_h} is the number of examinees who belong to the item response category k_h of item h and $\tilde{f}_{k_h}(\theta)$ is the estimated density. Asssume that the Old Test has a sufficiently large amount of test information at any point in the interval of θ of interest, and θ is transformed to τ, which has a constant test information throughout the corresponding interval of τ. Then

$$\tau_V = \tau + \varepsilon^* \ , \qquad [35]$$

where ε^* denotes the error in estimating $\hat{\tau}_V$.

In the Normal Approximation Method, a bivariate normal distribution is assumed for the joint distribution of $\hat{\tau}_V$ and τ for each subpopulation of examinees who share the same discrete item response k_h to the unknown item h. For simplicity, hereafter, the subscript V will be dropped from $\hat{\tau}_V$. The sample linear regression, $\dot{\varepsilon}^*(\hat{\tau})$ of the error ε^* on the maximum likelihood estimate $\hat{\tau}$ is given by

$$\dot{\varepsilon}^*(\hat{\tau}) = \alpha + \beta\hat{\tau} \ , \qquad [36]$$

where

$$\alpha = -C^{-2} E(\hat{\tau}) [Var(\hat{\tau})]^{-1} \qquad [37]$$

and

$$\beta = C^{-2} \left[\mathrm{Var}(\hat{\tau})\right]^{-1}] \ . \qquad [38]$$

For the conditional variance of ε^*, given $\hat{\tau}$,

$$\mathrm{Var}(\varepsilon^* | \hat{\tau}) \doteq C^{-2}[1 - C^{-2}\{\mathrm{Var}(\hat{\tau})\}^{-1}] \ . \qquad [39]$$

can also be written. Thus, the approximate joint distribution of ε^* and $\hat{\tau}$ are completely specified through Equations 36 through 39 for each subpopulation of examinees. Using the monte carlo method, as many error scores as desired can be produced for each value of $\hat{\tau}$; subtracting each value of ε^* thus produced from $\hat{\tau}$, the value of τ is obtained, which will be denoted by $\tilde{\tau}$. The resulting set of τ^* for each item response subgroup k_h will be appropriately categorized into small subintervals to provide the approximate product of the interval width of θ and the frequency $N_{k_h} \tilde{f}_{k_h}(\theta)$ in Equation 34, by virtue of the one-to-one correspondence between θ and τ; therefore, it can be used in both the numerator and denominator of Equation 34.

A hypothetical test, called Test A, which consists of 35 graded items, was used as the Old Test. Each item of Test A follows the normal ogive model on the graded response level, whose operating characteristics, $P_{x_g}(\theta)$, is given by Equation 35. The item discrimination parameter, $a_g (> 0)$, and the set of m_g (= 2) item difficulty parameters, b_{x_g}, of each of the 35 test items of Test A are presented in Table 3. The amount of test information of Test A is approximately 21.63 throughout the interval of θ (-3.0, 3.0); therefore, the standard error of estimate is approximately .215 for this interval of θ (cf. Samejima, 1977c). Five hundred hypothetical examinees whose positions on the latent trait θ are at 100 equally spaced points of θ starting with -2.475 and ending with 2.475, with five examinees placed at each position, were used. For Test A, τ can be set equal to θ, and the transformation of the latent trait is not necessary. These are 10 "unknown" binary items, each of which follows the normal ogive model on the dichotomous response level (whose item characteristic function is given by Equation 23, and whose two item parameters, a_g and b_g, are presented in Table 4).

Figure 1 presents the results obtained by the Normal Approximation Method (Samejima, 1977c), choosing Item 6 as an example. The result obtained by producing only one value of $\tilde{\theta}$ for each $\tilde{\theta}_v$ is plotted by hollow circles (N = 500); the two sets of results obtained by producing five values of $\tilde{\theta}$ for each are plotted by solid triangles (N = 2,500) and X's (N = 2,500), respectively; the result obtained by combining these two sets of five values of $\tilde{\theta}$ for each $\tilde{\theta}$ is plotted by hollow squares (N = 5,000). It can be seen that the accuracy of the estimation of the operating characteristic becomes higher as a larger number of $\tilde{\theta}$ are produced. Similar results were obtained for the other nine unknown test items, i.e., Items 1 through 5 and 7 through 10.

The Normal Approximation Method approximates the joint distribution of τ and $\hat{\tau}$ for each subpopulation by a bivariate normal distribution. A similar approach is possible with somewhat different rationale using the marginal density

Table 3
Item Discrimination Parameter a_g and Two Item Difficulty Parameters b_{xg} for x_g=1 and x_g=2 of Each of the 35 Graded Items of Test A

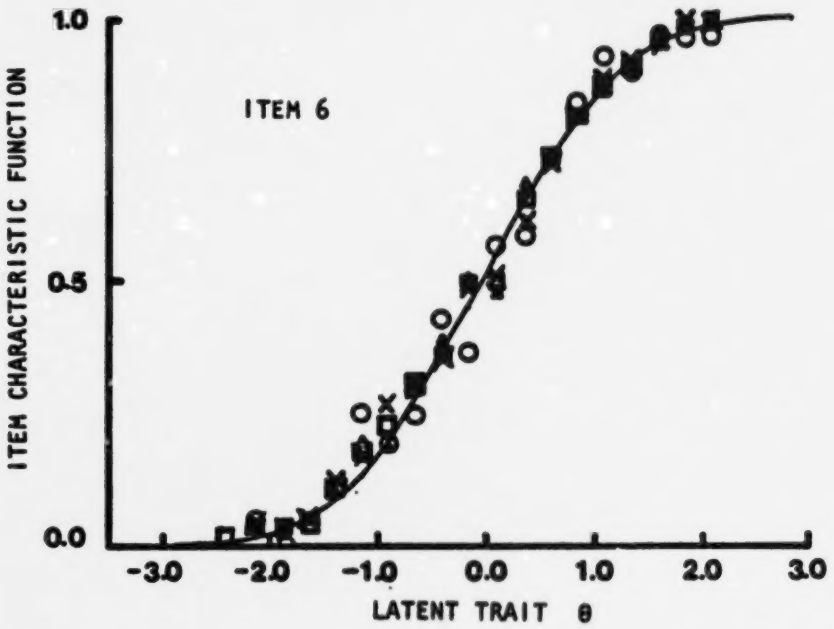
Item g	a_g	b_1	b_2
1	1.8	-4.75	-3.75
2	1.9	-4.50	-3.50
3	2.0	-4.25	-3.25
4	1.5	-4.00	-3.00
5	1.6	-3.75	-2.75
6	1.4	-3.50	-2.50
7	1.9	-3.00	-2.00
8	1.8	-3.00	-2.00
9	1.6	-2.75	-1.75
10	2.0	-2.50	-1.50
11	1.5	-2.25	-1.25
12	1.7	-2.00	-1.00
13	1.5	-1.75	-0.75
14	1.4	-1.50	-0.50
15	2.0	-1.25	-0.25
16	1.6	-1.00	0.00
17	1.8	-0.75	0.25
18	1.7	-0.50	0.50
19	1.9	-0.25	0.75
20	1.7	0.00	1.00
21	1.5	0.25	1.25
22	1.8	0.50	1.50
23	1.4	0.75	1.75
24	1.9	1.00	2.00
25	2.0	1.25	2.25
26	1.6	1.50	2.50
27	1.7	1.75	2.75
28	1.4	2.00	3.00
29	1.9	2.25	3.25
30	1.6	2.50	3.50
31	1.5	2.75	3.75
32	1.7	3.00	4.00
33	1.8	3.25	4.25
34	2.0	3.50	4.50
35	1.4	3.75	4.75

function of $\hat{\tau}$ (Samejima, 1978f). Let $g_{k_h}(\hat{\tau})$ denote the density function of the maximum likelihood estimate $\hat{\tau}$ and $\phi_{k_h}(\tau|\hat{\tau})$ be the conditional density function of τ, given $\hat{\tau}$, for each subpopulation of examinees who share the same discrete item response k_h. For the joint density function, $\xi_{k_h}(\tau,\hat{\tau})$, of τ and $\hat{\tau}$

Table 4
Item Discrimination Parameter a_h and Item Difficulty Parameter b_h of Each "Unknown" Test Item h

Item h	a	b
1	1.5	-2.5
2	1.0	-2.0
3	2.5	-1.5
4	1.0	-1.0
5	1.5	-0.5
6	1.0	0.0
7	2.0	0.5
8	1.0	1.0
9	2.0	1.5
10	1.0.	2.0

Figure 1
Estimated Item Characteristic Functions of Item 6 Based upon 500 $\tilde{\theta}$'s (Hollow Circles), 2,500 $\tilde{\theta}$'s (Solid Triangles), 2,500 $\tilde{\theta}$'s (Crosses), and 5,000's $\tilde{\theta}$'s (Hollow Squares), Obtained by the Normal Approximation Method, Using Test A as the Old Test



$$\xi_{k_h}(\tau,\hat{\tau}) = \phi_{k_h}(\tau|\hat{\tau})\ g_{k_h}(\hat{\tau}) \ . \qquad [40]$$

can be written. From these joint density functions for the separate subpopulations and the frequencies of examinees, N_{k_h}, are obtained for the estimated operating characteristic

$$\tilde{P}_{k_h}[\theta(\tau)] = N_{k_h}\int_{-\infty}^{\infty}\xi_{k_h}(\tau,\hat{\tau})d\hat{\tau}\ \left[\sum_j N_j\int_{-\infty}^{\infty}\xi_j(\tau,\hat{\tau})d\hat{\tau}\right]^{-1} \ . \qquad [41]$$

This approach is called the Bivariate P.D.F. Approach.

The density function, $g_{k_h}(\hat{\tau})$, in Equation 40 must be estimated from the observable set of maximum likelihood estimates. This can be done by fitting a polynomial of an appropriate degree to the set of maximum likelihood estimates by the method of moments (cf. Samejima & Livingston, 1979).

The conditional density, $\phi_{k_h}(\tau|\hat{\tau})$, in Equation 40 can be estimated by using the conditional moments of τ, given $\hat{\tau}$. It has been shown (Samejima, 1977e) that if an estimate of the latent trait is conditionally unbiased and its conditional distribution is normal with a constant second parameter, the conditional moments of the latent trait, given its maximum likelihood estimate, will be obtained solely from the density function of the estimate and the constant second parameter of the normal distribution. The conditional expectation of τ, given $\hat{\tau}$, and the second through fourth conditional moments of τ, given $\hat{\tau}$, are given by the following formulae:

$$E(\tau|\hat{\tau}) = \hat{\tau} + C^{-2}\frac{d}{d\hat{\tau}}\log g(\tau) \qquad [42]$$

$$\mathrm{Var}(\tau|\hat{\tau}) = C^{-2}\left[1 + C^{-2}\frac{d^2}{d\tau^2}\log g(\hat{\tau})\right] \qquad [43]$$

$$E[\{\tau - E(\tau|\hat{\tau})\}^3|\hat{\tau}] = C^{-6}\left[\frac{d^3}{d\tau^3}\log g(\hat{\tau})\right] \qquad [44]$$

$$E[\{\tau - E(\tau|\hat{\tau})\}^4|\hat{\tau}] = C^{-4}\left[3 + 6C^{-2}\left\{\frac{d^2}{d\tau^2}\log g(\hat{\tau})\right\} + 3C^{-4}\left\{\frac{d^2}{d\tau^2}\log g(\hat{\tau})\right\}^2 + C^{-4}\left\{\frac{d^4}{d\tau^4}\log g(\hat{\tau})\right\}\right] \qquad [45]$$

Let κ denote Pearson's criterion. Then,

$$\kappa = \beta_1(\beta_2 + 3)^2[4(2\beta_2 - 3\beta_1 - 6)(4\beta_2 - 3\beta_1)]^{-1} \qquad [46]$$

where

$$\beta_1 = (E[\{\tau - E(\tau|\hat{\tau})\}^3|\hat{\tau}])^2 (Var[\tau|\hat{\tau}])^{-3} \quad [47]$$

and

$$\beta_2 = (E[\{\tau - E(\tau|\hat{\tau})\}^4|\hat{\tau}])(Var[\tau|\hat{\tau}])^{-2} \ . \quad [48]$$

The value of this criterion points to a unique Pearson Type distribution (Elderton & Johnson, 1969) which approximates the conditional density, $\phi_{k_h}(\tau|\hat{\tau})$, in Equation 40. This method will be called the Pearson System Method (Samejima, 1978b).

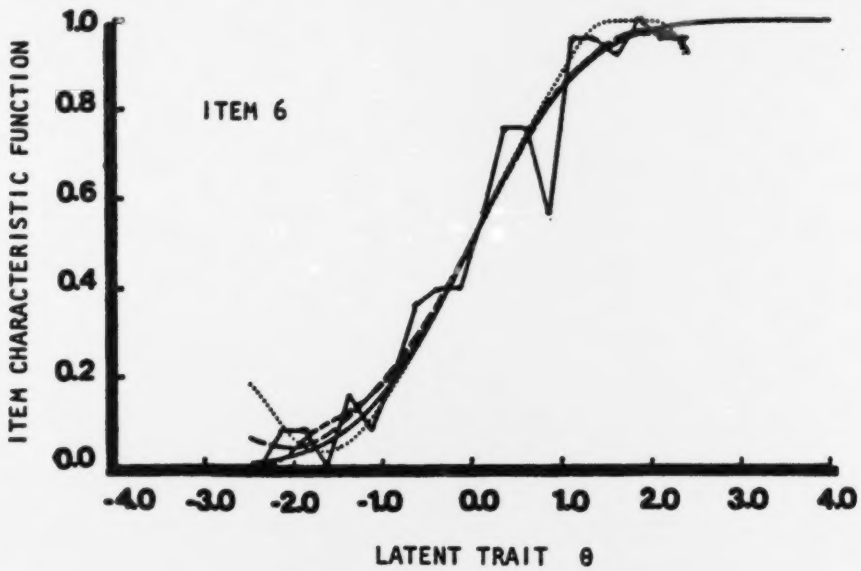
To avoid the increased inaccuracies of estimation of conditional moments of higher orders, it is sometimes suitable to use only the first and second conditional moments in approximating the conditional density, $\phi_{k_h}(\tau|\hat{\tau})$. The Normal Approach Method (Samejima, 1978b), which adopts a normal density function for the estimate of $\phi_{k_h}(\tau|\hat{\tau})$, is such a method, although each estimated conditional density function is forced to be symmetric. For added flexibility, a beta density function can also be used for the estimate of $\phi_{k_h}(\tau|\hat{\tau})$, whose two parameters, the lower and the upper endpoints of the interval for which the density is positive, are given a priori. Thus, as in the Normal Approach Method, only the first two conditional moments are needed to estimate the remaining two parameters of the beta distribution. This method will be called the Two-Parameter Beta Method (Samejima, 1977e, 1978a).

Figure 2 illustrates the results obtained by using the Bivariate P.D.F. Approach combined with the Normal Approach Method, with the same set of simulated data that was introduced earlier. The unknown test item illustrated here is, again, Item 6. Three different degrees, i.e., Degrees 3, 4 and 5, were used for the polynomials that were obtained by the method of moments to provide the estimated density function, $g_{k_h}(\hat{\theta})$, of each discrete item response k_h.

It can be seen in this example that the result of the Degree 3 case does not fit the theoretical item characteristic function as well as those of the Degree 4 and 5 cases, although it is still fairly close to the theoretical curve. This is not a general tendency, however, and those curves of the Degree 3 case fit just as closely to the corresponding theoretical item characteristic functions as those of the Degree 4 and 5 cases for most of the other nine unknown test items.

Comparison of these three resultant estimated item characteristic functions to the stepwise frequency ratios of the subgroup of examinees who answered Item 6 correctly to the total subgroup, which are based upon the true positions of the five hundred hypothetical examinees on the latent trait θ, indicates that all the three results are much closer to the theoretical item characteristic function than the frequency ratios.

Figure 2
Estimated Item Characteristic Functions of Item 6 for the Degree 3 (Dotted Curve), Degree 4 (Short Dashed Curve), and Degree 5 (Long Dashed Curve) Cases of the Bivariate P.D.F. Approach with the Normal Approach Method Using Test A as the Old Test, and the Theoretical Item Characteristic Function (Smooth Solid Curve) and the Frequency Ratios of θ (Jagged Solid Curve)



Conditional P.D.F. Approach

When there is a large number of unknown test items, the computing time required in the Normal Approximation Method or in the Bivariate P.D.F. Approach will be substantially large, since the joint density function must be approximated for each discrete item response to each unknown test item. In contrast, when it is desirable to deal with more than one unknown item together, the Conditional P.D.F. Approach is much simpler. This approach is further categorized into three procedures, i.e., the Simple Sum Procedure (Samejima, 1978a), the Weighted Sum Procedure (Samejima, 1978d), and the Proportioned Sum Procedure (Samejima, 1978e). Let $\hat{\tau}_s$ be the maximum likelihood estimate of the latent trait assigned to the examinee denoted $\underline{s}$, and $w(\hat{\tau}_s)$ denote an appropriately chosen weight for $\hat{\tau}_s$. In the Weighted Sum Procedure the estimated operating characteristic of the discrete item response k_h to the unknown test item $\underline{h}$ is given by

$$\tilde{P}_{k_h}[\theta(\tau)] = \sum_{s \epsilon k_h} w(\hat{\tau}_s)\, \tilde{\phi}(\tau|\hat{\tau}_s) \left[\sum_{s=1}^{N} w(\hat{\tau}_s)\, \tilde{\phi}(\tau|\hat{\tau}_s)\right]^{-1}, \qquad [49]$$

where $\phi(\tau|\hat{\tau}_s)$ is the estimated conditional density function of τ, given $\hat{\tau} = \hat{\tau}_s$.

In a special case where $w(\hat{\tau}_s) = 1$ for all $\hat{\tau}_s$'s, Equation 49 can be reduced to

$$\tilde{P}_{k_h}[\theta(\tau)] = \sum_{s \epsilon k_h} \tilde{\phi}(\tau|\hat{\tau}_s) \left[\sum_{s=1}^{N} \tilde{\phi}(\tau|\hat{\tau}_s) \right]^{-1} . \qquad [50]$$

The procedure which adopts Equation 50 instead of Equaton 49 is called the Simple Sum Procedure.

Let $p(s \epsilon k_h)$ be the probability with which the examinee $\underline{s}$ belongs to the item response subgroup k_h. This can be estimated by the proportion of examinees who belong to a specified discrete item response k_h. In the Proportioned Sum Procedure, for the estimated operating characteristic of the discrete item response k_h to the unknown test item $\underline{h}$

$$\tilde{P}_{k_h}[\theta(\tau)] = \sum_{s \epsilon k_h} \tilde{p}(s \epsilon k_h) \tilde{\phi}(\tau|\hat{\tau}_s) \left[\sum_{s=1}^{N} \tilde{\phi}(\tau|\hat{\tau}_s) \right]^{-1} \qquad [51]$$

can be written, where $\tilde{p}(s \epsilon k_h)$ is the proportion thus obtained as the estimate of the probability $p(s \epsilon k_h)$.

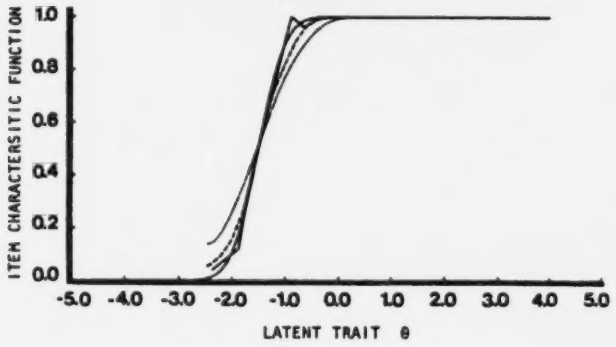
The conditional density, $\phi(\tau|\hat{\tau}_s)$, is approximated by the Pearson System Method, the Normal Approach Method, or the Two-Parameter Beta Method. In so doing, the conditional moments are obtained by Equations 42 through 45, with the replacement of the density function, $g_{k_h}(\hat{\tau})$, by $g(\hat{\tau})$, the density function of $\hat{\tau}$ for the total population of examinees. This is the primary advantage of the Conditional P.D.F. Approach, for the estimation of the marginal density function of $\hat{\tau}$ is done only for the total population, instead of for each item response subpopulation of for each unknown test item.

Figure 3 illustrates three examples obtained by the Simple Sum Procedure of the Conditional P.D.F. Approach, which is combined with the Normal Approach Method. The simulated data used in these examples are the same as those used in the examples given in the preceding section. The results are for the Degree 4 case, and similar results were also obtained for the Degree 3 case.

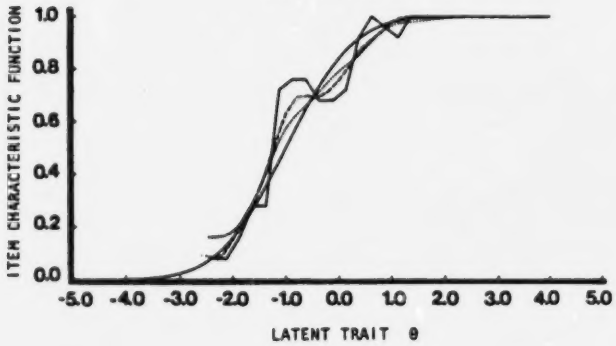
It can be seen, again, that the estimated item characteristic function is fairly close to the theoretical item characteristic function, and much more so in comparison to the corresponding frequency ratios, for each of the three unknown test items. In each of the three graphs, also presented is the resultant estimated item characteristic function of the Degree 4 case obtained by using a subtest of the original Old Test as the Old Test. The subtest consisted of 11 items of the original Old Test. Specifically, out of the 35 items, every 3rd item was chosen, starting with the 3rd and ending with the 33rd item (Samejima & Changas, 1981). Note that in this second situation, the original latent trait θ was transformed to τ, since the amount of test information function is not constant for the interval of θ, in which all the 500 hypothethical examinees were located. The resultant estimated item characteristic functions for the three unknown test items are fairly close to the corresponding theoretical item char-

Figure 3
Estimated Item Characteristic Functions Obtained by the Simple Sum Procedure of the Conditional P.D.F. Approach with the Normal Approach Method, Degree 4 Case, Using Test A (Dashed Curve) and Its Subtest of 11 Test Items (Dotted Curve) as the Old Test, Respectively, in Comparison with the Theoretical Item Charcteristic Function (Smooth Solid Curve) and the Frequency Ratios of the Latent Trait θ (Jagged Solid Curve)

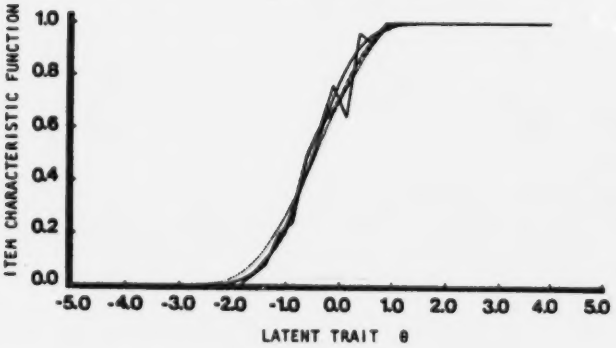
(a) Item 3



(b) Item 4



(c) Item 5



acteristic functions. For Item 4 the fit is even better than the one obtained using the original Old Test, although it is slightly poorer for Item 3 (Figure 3a) and approximately the same for Item 5 (Figure 3c). For each of the other seven unknown test items, the two sets of results turned out to be just as close to each other. That a reasonably good result can be obtained by using only 11 items in the Old Test indicates the robustness of the procedure. Similar studies have been made (Samejima, 1980b, 1981) using a subtest whose test items number 11 and 25. The results turned out to be just as good as those obtained using the original Old Test.

Computer Programs and Model Validation

Seven computer programs have been developed so far for the estimation of the operating characteristics of discrete item responses:

1. TAU TRANSFORMATION
2. MLE THETA
3. CONDITIONAL MOMENTS MLE
4. SIMPLE/WEIGHTED SUM NT
5. PROPORTIONED SUM NT
6. CONDITIONAL MOMENTS SUBGROUP
7. BIVARIATE P.D.F. NT

Figure 4 explains how to combine these programs in order to apply the Conditional P.D.F. Approach of the Bivariate P.D.F. Approach, which are combined either with the Normal Approach Method or with the Two-Parameter Beta Method. Each program is written in such a way that all the input data are printed separately from the results, and the user should pause and examine the result carefully and make some important decisions, if necessary, before proceeding to the next program.

Figure 4
Seven Programs for the Conditional P.D.F. Approach and for the Bivariate P.D.F. Approach, Each of Which Can be Combined Either with the Normal Approach Method or with the Two-Parameter Beta Method

Conditional P.D.F. Approach Combined with the Normal Approach or the Two-Parameter Beta Method:

TAU TRANSFORMATION MLE THETA	CONDITIONAL MOMENTS MLE	SIMPLE/WEIGHTED SUM NT PROPORTIONED SUM NT

Bivariate P.D.F. Approach Combined with the Normal Approach or the Two-Parameter Beta Method:

TAU TRANSFORMATION MLE THETA	CONDITIONAL MOMENTS SUBGROUP	BIVARIATE P.D.F. NT

Each program consists of several subroutines. There are three stages for each combination of programs. When the Conditional P.D.F. Approach is used, either SIMPLE/WEIGHTED SUM NT or PROPORTIONED SUM NT is chosen, depending upon whether the Simple Sum or Weighted Sum Procedure, or the Proportion Sum Procedure is used.

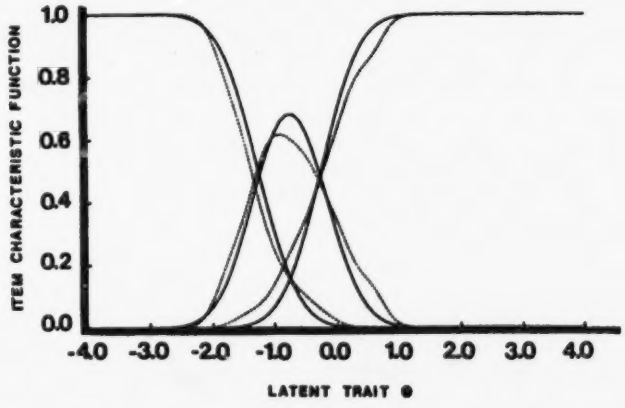
It is noted that the methods and approaches of estimating the operating characteristics of discrete test item responses introduced in the previous sections are directly applicable in such a situation where there is a validated item pool and it is desirable to add more test items to the item pool whose operating characteristics are yet to be discovered. When there is not such an item pool and it is necessary to start from the very beginning of test construction, however, there is no available set of test items which can be used as the Old Test. Thus, it may be necessary to assume some specific model, or models, for the discrete item responses of a set of test items, which were carefully selected with respect to their content validity, and to use them as the Old Test.

Notice that the same methods and approaches that have been developed for estimating the operating characteristics of the discrete item responses of unknown test items can be used for the purpose of model validation for the test items of the Old Test. If the adopted model, or models, is appropriate, then the estimated operating characteristics of the discrete item responses of each item of the Old Test should be close to those which are assumed by the model. This comparison can be performed by deleting a test item, using the remaining (n - 1) test items as the tentative Old Test, and estimating the operating characteristics of the discrete item responses of the deleted test item. If it is found that none of the resultant n sets of such estimated operating characteristics validates the adopted model, some other model will have to be used. If it is found that some, but not all, of the resultant n sets of estimated operating characteristics invalidate the adopted model, or models, then those test items may be excluded from the original Old Test, and the same process repeated for the remaining set of test items, and so forth. Research is still in progress using the original Old Test of 35 items which follow the normal ogive model with the parameters shown in Table 3.

For the purpose of illustration, Figure 5 presents the estimated operating characteristics of the three graded item responses of Item 15 of the original Old Test, which were obtained by the Simple Sum Procedure of the Conditional P.D.F. Approach combined with the Normal Approach Method, Degree 4 case, using the remaining 34 graded test items as the Old Test. The programs used in this process were Tau TRANSFORMATION, MLE THETA, CONDITIONAL MOMENTS MLE, AND SIMPLE/WEIGHTED SUM NT. It can be seen in this figure that each estimated operating characteristic is reasonably close to the corresponding theoretical operating characteristic.

The simple unweighted least squares method can be used, further, to evaluate the above result, by obtaining the estimates of the two parameters on the normal ogive model on the dichotomous response level, whose item characteristic function is given by Equation 23, as was done previously (Samejima, 1977c, 1977e). In the present example, the estimated operating characteristic, $\bar{P}_{x_g}(\theta)$,

Figure 5
Estimated Operating Characteristics for the Three Item Score Categories (Dotted Curves) of Item 15 of the Original Old Test, Using the Remaining 34 Graded Test Items as the Old Test in Comparison with the Theoretical Operating Characteristics (Solid Curves) Using the Simple Sum Procedure of the Conditional P.D.F. Approach with the Normal Approach Method



for $x_g = 1$, i.e., the intermediate item score, is added to the one for $x_g = 2$, to produce the estimated item characteristic function, which would be used if it were decided to restore the item dichotomously, assigning the binary item score 1 to $x_g = 1$ and $x_g = 2$ and 0 to $x_g = 0$. In a similar manner, the estimated operating characteristic $\tilde{P}_{x_g}(\theta)$ for $x_g = 2$ (i.e., the highest item score) can be treated as the estimated item characteristic function, which would be adopted if the item were rescored dichotomously, assigning 1 to $x_g = 2$ and 0 to $x_g = 1$ and $x_g = 0$. Thus, there are two estimated item characteristic functions for Item 15, having, presumably, the same discrimination parameter a_g and two different difficulty parameters b_g.

Let $\tilde{P}_g(\theta_j)$ be the estimated value of the item characteristic function in the normal ogive model for the midpoint of the jth interval, and define ζ_{gj} such that

$$\tilde{P}_g(\theta) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\zeta_{gj}} e^{-\frac{u^2}{2}} du \,. \qquad [52]$$

Following the least squares principle, Q is defined by

$$2Q = \sum_{j=1}^{m} [\zeta_{gj} - a_g(\theta_j - b_g)]^2 \,. \qquad [53]$$

Differentiating Q with respect to a_g and b_g and setting the results equal to zero gives

$$\frac{\partial Q}{\partial a_g} = \sum_{j=1}^{m} [\zeta_{gj} - a_g(\theta_j - b_g)](-\theta_j + b_g) = 0 \qquad [54]$$

and

$$\frac{\partial Q}{\partial b_g} = \sum_{j=1}^{m} [\zeta_{gj} - a_g(\theta_j - b_g)]\, a_g = 0 \,. \qquad [55]$$

The above equations provide the estimates of a_g and b_g such that

$$\hat{a}_g = \mathrm{Cov}(\zeta_{gj}, \theta_j)[\mathrm{Var}(\theta_j)]^{-1} \qquad [56]$$

and

$$\hat{b}_g = m_\theta - [\mathrm{Cov}(\zeta_{gj}, \theta_j)]^{-1}\mathrm{Var}(\theta_j)\bar{\zeta}_g \,, \qquad [57]$$

where m_θ and $\bar{\zeta}_g$ are the means of θ_j and ζ_{gj}, respectively. These estimates, $\hat{a}_g$ and $\hat{b}_g$, are to be compared with their respective parameters, a_g and b_g. The estimates of the two item parameters, $\hat{a}_g$ and $\hat{b}_g$, which are given by Equations 56 and 57, were obtained on each of the two estimated item characteristic functions. It should be noted that the estimated item difficulty parameter, $\hat{b}_g$, based on the first estimated item characteristic function is the estimate of b_{x_g} for $x_g = 1$, and that on the second estimated item characteristic function is the estimate of b_{x_g} for $x_g = 2$.

The estimation was made using the values of the estimated item characteristic function which are less than .9 and greater than .1, obtained for discrete values of θ with the step width of .1. In the first estimation, $\underline{m} = 15$ for the interval of θ, [-1.9, -0.5], and, in the second estimation, $\underline{m} = 18$ for the interval [-1.0, 0.7]. The two estimates of the discrimination parameter, a_g, were 1.790 and 1.421, respectively; these are somewhat less than the true parameter value 2.0 shown in Table 3, the result which is anticipated from Figure 5. The estimates of the two difficulty parameters, b_{x_g} for $x_g = 1$ and $x_g = 2$, were -1.302 and -0.230, respectively. It can be seen that they are reasonably close to the true parameter values, -1.250 and -0.250, given in Table 3.

Thus far the model used for the items of the Old Test has been restricted to the normal ogive model on the graded response level or on the dichotomous response level. Researchers have requested the development of methods of esti-

mating the operating characteristics of discrete item responses using the 3-parameter logistic model (Birnbaum, 1968) for the items of the Old Test or the set of test items whose operating characteristics are known. The item characteristic function, $P_g(\theta)$, of item g, which follows the 3-parameter logistic model is given by

$$P_g(\theta) = c_g + (1-c_g)\Psi_g(\theta) \ , \qquad [58]$$

where $\Psi_g(\theta)$ is the item characteristic function in the logistic model, for which

$$\Psi_g(\theta) = [1 + \exp\{-Da_g(\theta-b_g)\}]^{-1} \ , \qquad [59]$$

can be written and c_g is a constant that is supposed to be the probability of guessing correctly, or unity divided by the number of the alternatives attached to the multiple-choice item g.

There are several theoretical problems concerning the 3-parameter logistic model, including the fact that it does not assure the existence of the unique maximum likelihood estimate of the latent trait θ for every response pattern (cf. Samejima, 1972, 1973b). This is caused by the fact that the item response information function, $I_{k_g}(\theta)$, which is defined by Equation 5 for the discrete item response k_g, assumes negative values for the positive response to the binary item g for the interval of θ less than the critical value θ_g (cf. Samejima, 1973b). This critical value is considerably high, relative to the difficulty parameter b_g. For example, when $D = 1.7$, $a_g = 1.0$, $b_g = 0.0$, and $c_g = .20$, the critical value θ_g is as high as .473. If the value of c_g is changed from .20 to .25, i.e., the item is changed from a five-alternative multiple-choice test item to a four-alternative one, this gives $\theta_g \doteq -0.408$.

These facts indicate that a test item which follows the 3-parameter logistic model has a serious problem caused by the random guessing noise. In fact, in the above example, $P_g(\theta) = 0.5$ at $\theta = \theta_g$, which can be partitioned into two parts, i.e., .333 for knowledge and .167 for guessing correctly. This indicates that the effect of noise is as large as 50% of that of the true information. The use of the test items following the 3-parameter logistic model for the Old Test has this serious problem, and methods should be developed in such a way that test items of lower levels relative to the individual examinee's ability level must be excluded from the maximum likelihood estimation of his or her ability. It can be forseen that the method will necessarily be much more complicated than those in which a model such as the normal ogive model or the logistic model is used.

Information Loss Caused by Asymptotes

It has been seen in the preceding section that the 3-parameter logistic model, in which the item characteristic function has a lower asymptote greater than zero, has the interval, $(-\infty, \theta_g)$, for which the item response information function for the positive response assumes negative values. Even in such a situation, the item information function, $I_g(\theta)$, is nonnegative throughout the en-

tire range of θ, as is obvious from Equation 6 (see Table 2). It is questionable, however, if the item information function is just as meaningful in such a situation as it is when all the item response information functions are nonnegative.

For the moment, the above question will be kept open, and investigation will continue into the amount of the loss of information caused by the lower asymptote which is greater than zero, or by the upper asymptote which is less than unity. In so doing, there will only be consideration of item characteristic functions that are strictly increasing in θ, with c_{g1} and c_{g2} as the lower and upper asymptotes, respectively. Thus,

$$\begin{cases} \lim_{\theta \to \underline{\theta}} P_g(\theta) = c_{g1} , \\ \lim_{\theta \to \bar{\theta}} P_g(\theta) = c_{g2} \end{cases} \qquad [60]$$

where $\underline{\theta}$ and $\bar{\theta}$ are the lower and upper endpoints of the interval for which θ is defined, and

$$0 \leq c_{g1} < c_{g2} \leq 1 . \qquad [61]$$

Let $\underline{j}$ be a binary test item with $c_{j1} = 0$ and $c_{j2} = 1$. Define τ such that

$$\tau = \tau(\theta) = P_j^{-1}[P_g(\theta)] . \qquad [62]$$

Thus, τ is a strictly increasing function of θ with the range

$$\underline{\tau} < \tau < \bar{\tau} , \qquad [63]$$

where

$$\begin{cases} \underline{\tau} = P_j^{-1}[P_g(\underline{\theta})] = P_j^{-1}(c_{g1}) \\ \bar{\tau} = P_j^{-1}[P_g(\bar{\theta})] = P_j^{-1}(c_{g2}) . \end{cases} \qquad [64]$$

For the two item information functions, $I_g(\theta)$ and $I_j^*(\tau)$,

$$\int_{\underline{\theta}}^{\bar{\theta}} [I_g(\theta)]^{\frac{1}{2}} \, d\theta = \int_{\underline{\tau}}^{\bar{\tau}} [I_j^*(\tau)]^{\frac{1}{2}} \, d\tau . \qquad [65]$$

can be written. For convenience, and without loss of generality, let $\underline{j}$ follow the logistic model, such that

$$P_j(\tau) = [1 + \exp\{-Da_j(\tau - b_j)\}]^{-1} . \qquad [66]$$

Thus,

$$\begin{cases} \underline{\tau} = [Da_j]^{-1}[\log c_{g1} - \log(1-c_{g1})] + b_j \\ \bar{\tau} = [Da_j]^{-1}[\log c_{g2} - \log(1-c_{g2})] + b_j \end{cases} . \qquad [67]$$

From Equation 67 it can be seen that if $c_{g1} = 0$, then $\underline{\tau} = -\infty$, otherwise, $\underline{\tau}$ is finite; and if $c_{g2} = 1$, then $\bar{\tau} = \infty$, otherwise, $\bar{\tau}$ is finite. From Equation 65 and 66 for the total information Q is obtained:

$$Q = \int_{\underline{\theta}}^{\bar{\theta}} [I_g(\theta)]^{1/2} \, d\theta$$

$$= Da_j \int_{\underline{\tau}}^{\bar{\tau}} [\exp\{Da_j(\tau-b_j)\}]^{1/2} [1 + \exp\{Da_j(\tau-b_j)\}]^{-1} \, d\tau . \qquad [68]$$

Now τ^* will be further defined such that

$$\tau^* = \tau^*(\tau) = [\exp\{Da_j(\tau-b_j)\}]^{1/2} . \qquad [69]$$

Thus,

$$\frac{d\tau}{d\tau^*} = 2(Da_j\tau^*)^{-1} , \qquad [70]$$

and from Equations 67 and 69

$$\underline{\tau}^* = \tau^*(\underline{\tau}) = c_{g1}^{1/2} (1-c_{g1})^{-1/2} \qquad [71]$$

and

$$\bar{\tau}^* = \tau^*(\bar{\tau}) = c_{g2}^{1/2} (1-c_{g2})^{-1/2} . \qquad [72]$$

can be written. From Equations 68, 69, 70, 71, and 72 is obtained

$$Q = 2[\tan^{-1}\{c_{g2}/(1-c_{g2})\}^{1/2} - \tan^{-1}\{c_{g1}/(1-c_{g1})\}^{1/2}] . \qquad [73]$$

It is obvious from Equation 73 that when $c_{g1} = 0$ and $c_{g2} = 1$, as is the case with the normal ogive model or with the logistic model on the dichotomous response level, this quantity, Q, assumes the maximum value, such that

$$Q = \pi . \qquad [74]$$

It can also be seen from Equation 73 that as c_{g1} departs from zero and c_{g2} from unity, the total information Q becomes progressively smaller than π.

When item g follows the 3-parameter logistic model, or 3-parameter normal ogive model, this gives

$$\begin{cases} c_{g1} = c_g > 0 \\ c_{g2} = 1 \ . \end{cases} \qquad [75]$$

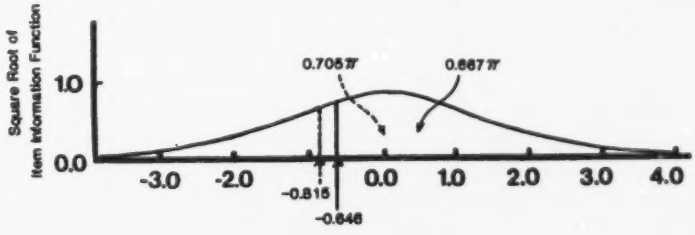
From Equation 73 and 74 can be written

$$Q = \pi - 2\tan^{-1}[c_g/(1-c_g)]^{1/2} \ . \qquad [76]$$

Thus, it can be seen that the second term in the right-hand side of Equation 76 indicates the information loss caused by the asymptote c_g, which is greater than zero. This information loss is as large as $.295\pi$ for $c_g = .20$ (i.e., when the multiple-choice test item has five alternative answers); when $c_g = .25$ (i.e., the multiple-choice test item has four alternative answers), it is as large as $.333\pi$, which indicates that one-third of the total information is lost.

Figure 6 illustrates the information loss caused by the lower asymptote, which is greater than zero. In this example, the solid curve indicates the square root of the item information function in the logistic model, with D = 1.7, $a_j = 1.0$, and $b_j = 0.0$. The total area under the solid curve above the abscissa equals π, the left-hand side of the vertical dashes indicates the information loss caused by $c_j = .20$, and the left-hand side of the vertical solid line corresponds to the information loss caused by $c_j = .25$.

Figure 6
Information Loss Caused by the Lower Asymptote c_g of the Item Characteristic Function in the 3-Parameter Logistic Model for $c_g = .20$ (Dashes) and for $c_g = 0.25$ (Solid Line)



The information loss caused by the lower asymptote c_g was computed for each of the three additional values of c_g, i.e., .10, .33, and .50; they are all pre-

sented in Table 5 together with the corresponding values of τ. In the same table, also presented are the critical value θ_g; the values of $\Psi_g(\theta)$ and $P_g(\theta)$; and that of the item information function, $I_g(\theta)$, at $\theta = \theta_g$, for which

$$\theta_g = (2Da_g)^{-1} \log c_g + b_g \ , \qquad [77]$$

$$\Psi_g(\theta_g) = [1 + \exp\{-Da_g(\theta_g - b_g)\}]^{-1} = c_g^{\frac{1}{2}}(1+c_g^{\frac{1}{2}})^{-1} \ , \qquad [78]$$

$$P_g(\theta_g) = c_g + (1-c_g)\Psi_g(\theta_g) = c_g^{\frac{1}{2}} \ . \qquad [79]$$

and

$$I_g(\theta_g) = (1-c_g)D^2a_g^2\Psi_g(\theta_g)^2[1-\Psi_g(\theta_g)][c_g+(1-c_g)\Psi_g(\theta_g)]^{-1}$$

$$= D^2a_g^2c_g^{\frac{1}{2}}(1-c_g^{\frac{1}{2}})(1+c_g^{\frac{1}{2}})^{-2} \ . \qquad [80]$$

can be written.

Table 5
Information Loss Caused by the Lower Asymptote of the Item Characteristic Function in the 3-Parameter Logistic Model for Each of the Five Values of the Asymptote c_g Together with τ, the Critical Value θ_g, and the Logistic Function $\Psi_g(\theta)$, the Item Characteristic Function $P_g(\theta)$ and the Item Information Function $I_g(\theta)$ at $\theta = \theta_g$

c_g	τ	Information Loss	θ_g	$\Psi_g(\theta_g)$	$P_g(\theta_g)$	$I_g(\theta_g)$
.10	-1.292	.644 (20.5%)	.677	.240	.316	.361
.20	.815	.927 (29.5%)	.473	.309	.447	.341
.25	.646	1.047 (33.3%)	.408	.333	.500	.321
.33	.408	1.231 (39.2%)	.323	.366	.577	.283
.50	.000	1.571 (50.0%)	.204	.414	.707	.205

Shiba's Word/Phrase Comprehension Tests and Scale Construction

The 13 word/phrase comprehension tests developed by Shiba (1978), which were introduced by Samejima (1980d), are based upon an item pool of 480 test items. Out of those 13 tests, tests AP1 and AP2 are for young children who have not yet learned how to read, and pictures are used instead of words and phrases. Tests A1, A2, A3, A4, A5, nd A6 are basicially for elementary school children of six different grades. Two tests, J1 and J2, are basically for junior high school students of three different grades; and Tests S1 and S2 are for senior

high school students of three different grades. Finally, Test U2 is for college students and adults. Each test item is a multiple-choice item, with five alternative answers, i.e., one correct answer and four distractors. The distractors were carefully chosen in such a way that the choice of each alternative will provide information, as does the choice of the correct answer.

Shiba's item pool of 480 items is constructed so that a certain number of test items are used in two adjacent tests; Table 6 presents the total number of test items in each test, together with the number of test items which are shared by each pair of adjacent tests. As can be seen, the numbers of test items for separate tests vary between 30 and 60, and the numbers of shared test items are between 8 and 20.

Table 6
Number of Items of Each of Shiba's 13 Tests of Words/Phrase Comprehension, and Number of Shared Items for Each Pair of Tests

Test	Number of Items	No. of Shared Items
AP1	30	
		10
AP2	30	
		8
A1	36	
		8
A2	34	
		16
A3	40	
		8
A4	40	
		16
A5	48	
		16
A6	54	
		16
J1	56	
		20
J2	60	
		20
S1	60	
		20
S2	60	
		20
U2	57	

When an item pool of a broad range of difficulty levels, such as Shiba's, is available, a number of different uses can be made of it, which may lead to fruitful research results. In so doing, scale construction with a sound theoretical background is most desirable. If this turns out to be successful, then it will be possible to compare, for instance, the ability of two individuals who differ substantially in age and education, to conduct developmental research by following up single individuals, and so forth.

As the first trial, Shiba and others factor analyzed each of 11 of the tests (excluding AP1 and AP2) separately, using a set of data obtained upon a single group of examinees, and then combined the results by equating the item parameters of overlapping test items. The normal ogive model, or the logistic model as its approximation, was assumed for each multiple-choice test item which was scored dichotomously. They were aware that the general factor, which they found for each set of a test and a group of examinees, might not be the same general factor they found for another set, and it is an assumption that all those general factors are one single factor.

To ameliorate this relative weakness, it is conceivable to apply simultaneous factor analysis for each subset of test items which is shared by two adjacent tests, using two or more different groups of examinees. If it turns out that a general factor exists across two or more groups of examinees, then it will be stronger support for the unidimensionality of the latent trait.

In addition to the above benefit, there is a strong possibility that a subset of overlapping test items can be used as the Old Test, from which can be estimated the operating characteristics of both the correct answers and distractors of the other nonoverlapping test items of the two adjacent tests (Shiba, Noguchi, & Haebara, 1978). If the resultant estimated operating characteristics turn out to be similar to the plausibility functions provided by the family of models developed for the multiple-choice item (Samejima, 1979b), then use can be made of the information given by the distractors, as well as the information provided by the correct answer, in estimating the examinee's latent trait. The same process can be also applied for estimating the operating characteristics of the distractors of the test items of the Old Test itself, as well as the model validation for their correct answers, in a similar manner which was illustrated in the sixth section.

The estimation of the operating characteristics will be more accurate if a test has two subsets of shared test items, as is the case with most of Shiba's tests, on lower and upper sides of the latent trait. Through the results of this estimation, it will be possible to equate the scales more rigorously, not only through the sets of shared subsets of test items but through all the other test items.

In applying simultaneous factor analysis, it is necessary that there are covariance matrices, instead of correlation matrices, for what is needed is a common scale for each item variable across the two or more populations of examinees involved. In a situation such as this, however, the covariance matrices are not readily available, since the item variable, which is denoted by X_g, is a

hypothesized variable behind the item score. On the other hand, with a normal distribution assumption, the tetrachoric correlation matrix can be obtained for each group of examinees. A way must be found, therefore, to convert tetrachoric correlation matrices for two or more populations to covariance matrices.

There is a simple, straightforward solution for the above problem when a distractor or distractors deserving the second best item score can be located. Here, a simple case will be considered in which there are only two populations of examinees, to whom a common subtest, including item g, was administered. It is assumed that the item variable, X_g, distributes normally for each of the two populations of examinees, with μ_{g1} and σ_{g1} as the two parameters for Populations 1 and μ_{g2} and σ_{g2} for Populations 2. Let p_{gr1} and p_{gr2} be the probabilities with which the examinee chooses the alternative with the rth highest item score for item g, for Populations 1 and 2, respectively, and γ_{gr1} and γ_{gr2} denote the corresponding normal deviates, such that

$$\begin{cases} \gamma_{gr1} = \Phi^{-1}[1 - \sum_{s=1}^{r} p_{gs1}] \\ \gamma_{gr2} = \Phi^{-1}[1 - \sum_{s=1}^{r} p_{gs2}] \end{cases} , \qquad [81]$$

where Φ indicates the standard normal distribution function. Thus, for the ratio of the two standard deviations

$$\sigma_{g2}/\sigma_{g1} = (\gamma_{gr1} - \gamma_{gs1})/(\gamma_{gr2} - \gamma_{gs2}) . \qquad [82]$$

can be written. When one set of distractors which deserves the second best item score can be located, r = 2 can be set for the correct answer and s = 1 for the distractors, and an estimate of the ratio of the standard deviations of the item variable X_g can be obtained for Populations 1 and 2 through Equation 81, by using the corresponding sample proportions for p_{gr1}, p_{gr2}, p_{gs1}, and p_{gs2}. When, in addition, another set of distractors which deserves the third best item score, and so on, can be located, there is more than one set of relationships described by Equation 81. In such a case, some least square procedure will provide an estimate of the ratio of the two standard deviations. In so doing, it is desirable to allow errors on both sides such as is done by Deming's (1946) least squares method. For the distance between the two means of X_g

$$\mu_{g2} - \mu_{g1} = \sigma_{g1}\gamma_{gr1} - \sigma_{g2}\gamma_{gr2} . \qquad [83]$$

can be written. Since there is more than one value of r in Equation 83, the arithmetic mean of the results of Equation 83 will provide an estimate of the distance between the two means.

Table 7 presents two contingency tables for the choice of the alternatives of Item 51 of Shiba's Test J1 against five test score groups, for the samples of seventh graders and eighth graders, respectively. It can be seen in these two tables that, next to the correct answer E, Alternative C attracted examinees of

Table 7
Contingency Tables of the Five Alternatives--A, B, C, D, and E--of Item 51 of Test J1 vs. Five Test Score Categories Together with the Mean of Maximum Likelihood Estimate of θ for Each Alternative Subgroup for Seventh- and Eighth-Grade Students

Subgroup and Test Score	Alternative					No Answer	Total
	A	B	C	D	E*		
Seventh Grade							
Lowest	27	36	11	30	22	4	130
Low	10	20	12	22	0	3	107
Middle	23	10	25	28	32	2	120
High	21	16	18	24	33	6	118
Highest	22	4	12	21	79	1	139
Total	103	86	78	125	206	16	614
Mean $\hat{\theta}$	.137	.640	.077	.136	.429	—	
Eighth Grade							
Lowest	25	18	10	17	12	3	85
Low	13	14	9	16	27	2	81
Middle	17	8	14	16	48	5	108
High	10	6	18	13	38	2	87
Highest	11	1	4	6	77	1	100
Total	76	47	55	68	202	13	461
Mean $\hat{\theta}$	.118	.260	.312	.150	.909	—	

*Correct Answer

Figure 7
Relationship among the Proportions of Examinees Who Correctly Answered Item 4 of Test J1 and Who Selected the Second Best Alternative, and the Means and Standard Deviations of the Two Populations JH1 and JH2

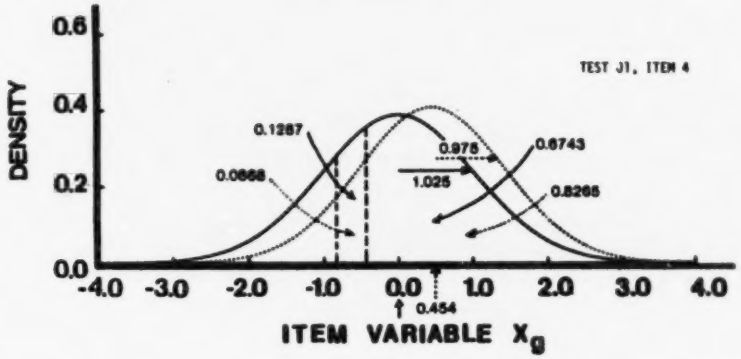


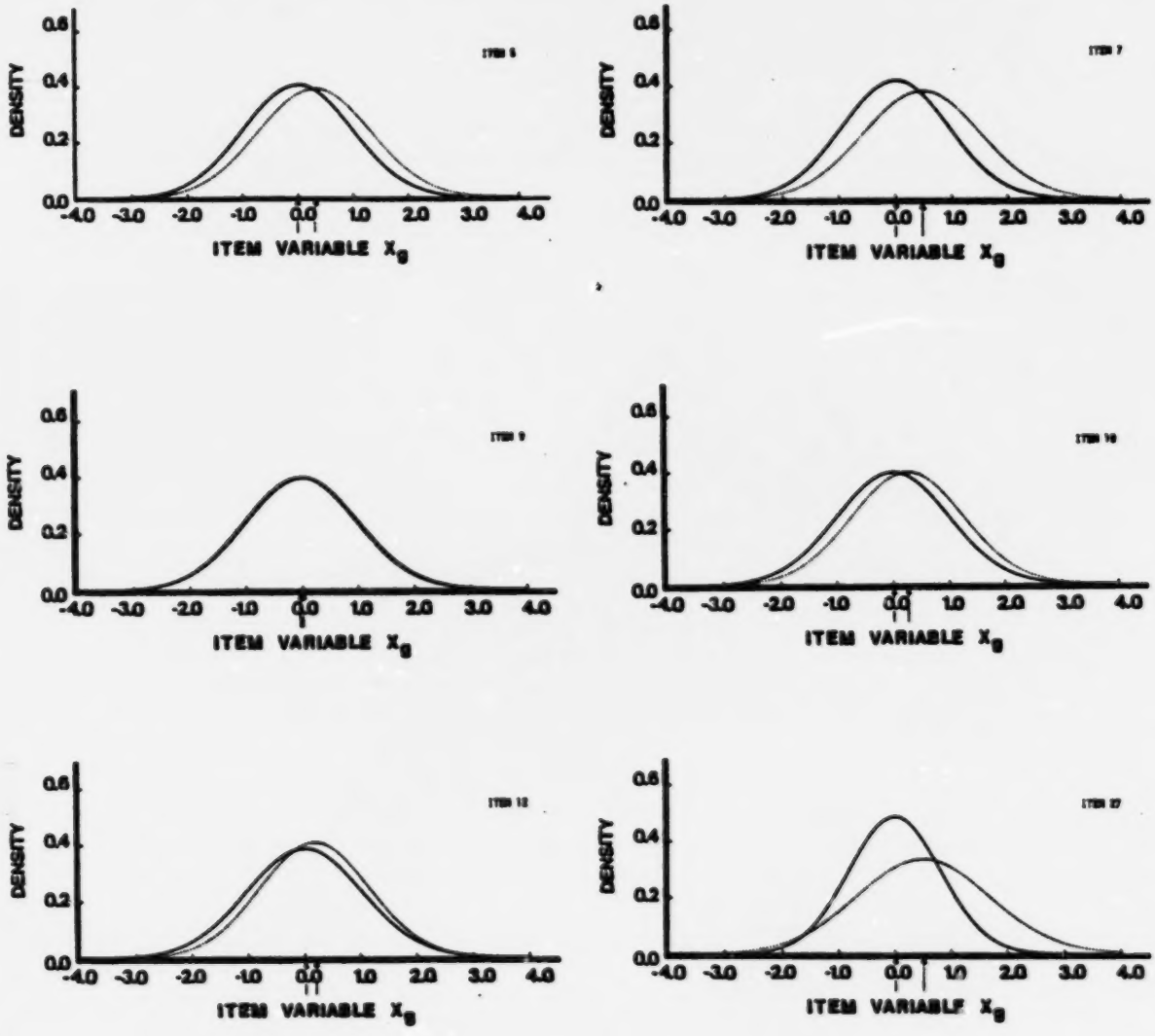
Table 8
Estimated Means and Standard Deviations of Each of the 14 Item Variables of Test J1 for the Two Populations, JH1 and JH2, Together with the Correct Answer and Alternative(s) of the Second Best Item Score

Item	Mean		SD		Correct Answer	Second Best
	JH1	JH2	JH1	JH2		
4	.000	.454	1.025	.975	E	A
5	.000	.308	.982	1.018	A	D, E
7	.000	.479	.953	1.047	D	B
9	.000	-.047	1.002	.998	C	D, E
10	.000	.266	1.000	1.000	E	B
12	.000	.185	1.026	.974	B	A, E
27	.000	.499	.816	1.184	E	C
29	.000	.131	1.017	.983	C	D
30	.000	.130	.973	1.027	B	A, D
34	.000	.162	.945	1.055	B	D
47	.000	.743	.730	1.270	B	C, E
48	.000	.131	1.008	.992	E	A, B
50	.000	.106	1.037	.963	A	B, D
51	.000	.241	.952	1.048	E	C

Table 9
Factor Loadings of First Three Common Factors Obtained by Simultaneous Factor Analysis of 14 Items of Test J1 Using JH1 and JH2 as Two Populations

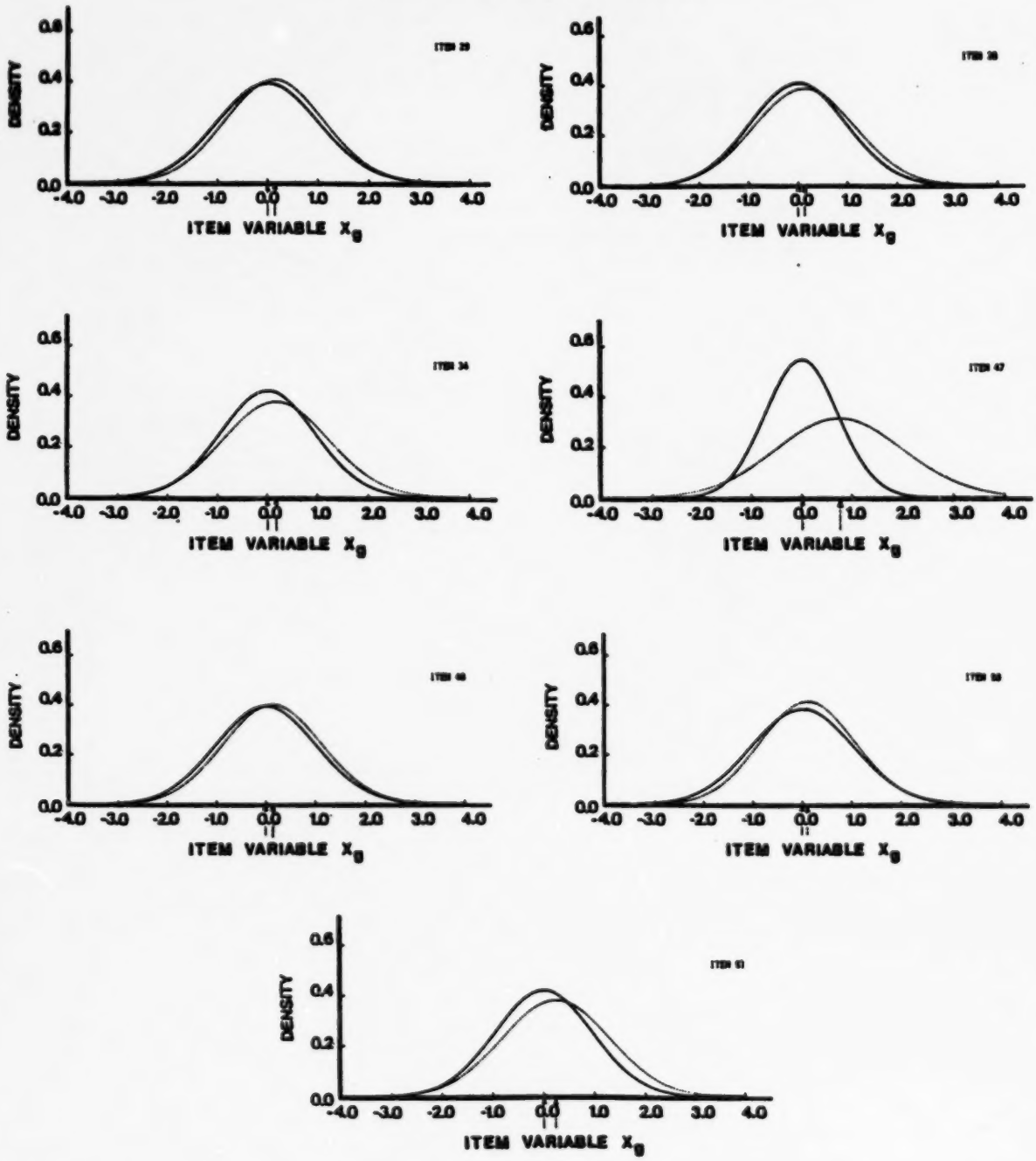
Item No.	Common Factor		
	1	2	3
4	.615	.099	.175
5	.597	-.056	-.178
7	.405	-.330	-.161
9	.340	-.069	-.040
10	.480	-.154	-.418
12	.446	.069	.147
27	.320	-.484	.127
29	.613	-.016	-.271
30	.308	-.070	-.151
34	.286	.048	.066
47	.332	.189	.073
48	.412	-.182	.124
50	.366	-.278	.037
51	.389	-.154	.350

Figure 8
Estimated Density Functions of Item Variable X_g for
the Seventh (Solid Curve) and Eighth (Dotted Curve) Graders
for Each of 13 Items of Test J1



- continued on the next page -

Figure 8 (Continued)
Estimated Density Functions of Item Variable X_g for
the Seventh (Solid Curve) and Eighth (Dotted Curve) Graders
for Each of 13 Items of Test J1



BEST COPY AVAILABLE

relatively high levels of ability, and Alternative D probably follows C in that tendency. This observation from the contingency tables is supported by each of the two configurations of the means of maximum likelihood estimates θ for the separate alternative subgroups of examinees (cf. Samejima, 1980d), which are given at the bottom of each contingency table. Thus either Alternative C alone, or the combination of Alternative C and D, should be treated as the one which deserves the second best item score.

Figure 7 presents the relationship between the two sets of parameters of the item variable distributions for Item 4 of Test J1, by setting $\mu_{g1} = 0$ and $\sigma_{g1} + \sigma_{g2} = 2$. Fourteen test items have been chosen out of the 56 of Test J1, for each of which the identification of the second best alternative answer, or answers, was made relatively easily. The estimated means and standard deviations for the two populations, JH1 and JH2, for each of the 14 test items, together with the correct answer and the alternative(s) with the second highest item score, are given in Table 8. Figure 8 also presents the estimated density functions of the item variable X_g for the two populations for each of the remaining 13 test items of Test J1, excluding Item 4, which was already illustrated in Figure 7. It can be seen from the table and figures that, except for a few items, such as Items 27 and 47, the two estimated density functions are close to each other.

By multiplying the estimated tetrachoric correlation coefficient between the two item variables, X_g and X_j, with the product of the corresponding two estimated standard deviations, the correlation coefficient can be converted to the covariance of each pair of test items, g and j, and hence the sample correlation matrix to the sample covarince matrix for each population.

Table 10
Covariance Matrices of
the Three Common Factors
Resulting from Simultaneous
Factor Analysis for
Populations JH1 and JH2

Population and Common Factor	Common Factor 1	2	3
JH1			
1	1.000		
2	0.000	1.000	
3	0.000	0.000	1.000
JH2			
1	1.072		
2	-.324	-.091	
3	.354	.393	1.037

The factor loadings of the first three common factors for the 14 test items obtained by the simultaneous factor analysis (SIFASP) for the two examinee groups, JH1 and JH2, are presented in Table 9. They are orthogonal and standardized factors for JH1 and oblique for JH2. The covariance of these three factors for JH1 and JH2 are given as Table 10. It can be seen from Table 9 that the first common factor can be considered as the general factor for these 14 item variables.

REFERENCES

Birnbaum, A. Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick, Statistical theories of mental test scores. Reading MA: Addison-Wesley, 1968.

Bock, R. D. Estimating item parameters and latent ability when responses are scored in two or more nominal categories. Psychometrika, 1972, 37, 29-51.

Deming, W. E. Statistical adjustment of data. New York: Wiley, 1946.

Elderton, W. P., & Johnson, N. L. Systems of frequency curves. Cambridge England: Cambridge University Press, 1969.

Kendall, M. G., & Stuart, A. The advanced theory of statistics (Vol. 2). New York: Hafner, 1961.

Lawley, D. N., & Maxwell, A. E. Factor analysis as a statistical method. London: Butterworth, 1971.

Levine, M. Appropriateness measurement and the formula-score method: Overview, intercorrelations and interpretations. Paper presented at the Office of Naval Research conference on model-based psychological measurement, Iowa City IA, 1980.

Lord, F. M. A theory of test scores. Psychometric Monograph, No. 7, 1952.

Lord, F. M. Estimating true-score distributions in psychological testing (an empirical Bayes estimation problem). Psychometrika, 1969, 34, 259-299.

Lord, F. M. Item characteristic curves estimated without knowledge of their mathematical form--A confrontation of Birnbaum's logistic model. Psychometrika, 1970, 35, 43-50.

Lord, F. M., & Novick, M. R. Statistical theories of mental test scores. Reading MA: Addison-Wesley, 1968.

Rasch, G. Probabilistic models for some intelligence and attainment tests (expanded edition). Chicago: University of Chicago Press, 1980. (Originally published, Copenhagen: Danmarks Paedagogiske Institut, 1960)

Samejima, F. Estimation of ability using a response pattern of graded scores. Psychometrika Monograph, No. 17, 1969.

Samejima, F. A general model for free-response data. Psychometrika Monograph, No. 18, 1972.

Samejima, F. Homogeneous case of the continuous response level. Psychometrika, 1973, 38, 203-219. (a)

Samejima, F. A comment on Birnbaum's three-parameter logistic model in the latent trait theory. Psychometrika, 1973, 38, 221-233. (b)

Samejima, F. Normal ogive model on the continuous response level in the multidimensional latent space. Psychometrika, 1974, 39, 111-121. (a)

Samejima, F. Normal ogive model on the graded response level in the multidimensional latent space. Paper presented at the annual meeting of the Psycometric Society, Stanford CA, April 1974. (b)

Samejima, F. Graded response model of the latent trait theory and tailored testing. In C. L. Clark (Ed.), Proceedings of the first conference on computerized adaptive testing (U.S. Civil Service Commission, Personnel Research and Development Center, PS-75-6). Washington DC: U.S. Government Printing Office, 1976. (Superintendent of Documents Stock No. 006-00940-9)

Samejima, F. Effects of individual optimization in setting the boundaries of dichotomous items on accuracy of estimation. Applied Psychological Measurement, 1977, 1, 77-94. (a)

Samejima, F. A use of the information function in tailored testing. Applied Psychological Measurement, 1977, 1, 233-247. (b)

Samejima, F. A method of estimating item characteristic functions using the maximum likelihood estimate of ability. Psychometrika, 1977, 42, 163-191. (c)

Samejima, F. Weakly parallel tests in latent trait theory with some criticisms of classical test theory. Psychometrika, 1977, 42, 193-198. (d)

Samejima, F. Estimation of the operating characteristics of item response categories I: Introduction to the two-parameter beta method (Research Report 77-1). Knoxville: University of Tennessee, Department of Psychology, 1977. (e)

Samejima, F. Estimation of the operating characteristics of item response categories II: Further development of the two-parameter beta method (Research Report 78-1). Knoxville: University of Tennessee, Department of Psychology, December 1978. (a)

Samejima, F. Estimation of the operating characteristics of item response categories III: The normal approach method and the Pearson system method (Re-

search Report 78-2). Knoxville: University of Tennessee, Department of Psychology, June 1978. (b)

Samejima, F. Estimation of the operating characteristics of item response categories IV: Comparison of the different methods (Research Report 78-3). Knoxville: University of Tennessee, Department of Psychology, June 1978. (c)

Samejima, F. Estimation of the oeprating characteristics of item response categories V: Weighted sum procedure in the conditional P.D.F. approach (Research Report 78-4). Knoxville: University of Tennessee, 1978. (d)

Samejima, F. Estimation of the operating characteristics of item response categories VI: Proportioned sum procedure in the conditional P.D.F. approach (Research Report 78-5). Knoxville: University of Tennessee, Department of Psychology, 1978. (e)

Samejima, F. Estimation of the operating characteristics of item response categories VII: Bivariate P. D. F. approach with normal approach method (Research Report 78-6). Knoxville: University of Tennessee, Department of Psychology, December 1978. (f)

Samejima, F. Constant information model: A new, promising item characteristic function (Research Report 79-1). Knoxville: University of Tennessee, Department of Psychology, February 1979. (a)

Samejima, F. A new family of models for the multiple-choice item (Research Report 79-4). Knoxville: University of Tennessee, Department of Psychology, December 1979. (b)

Samejima, F. Estimation of the operating characteristics when the test information of the old test is not constant I: Rationale (Research Report 80-2). Knoxville: University of Tennessee, Department of Psychology, June 1980. (a)

Samejima, F. Estimation of the operating characteristics when the test information of the old test is not constant II: Simple sum procedure of the conditional P. D. F. approach/normal approach method using three subtests of the old test (Research Report 80-4). Knoxville: University of Tennessee, Department of Psychology, November 1980. (b)

Samejima, F. Latent trait theory and its applications. In P. R. Krishnaiah (Ed.), Multivariate Analysis V. Amsterdam, Netherlands: North-Holland, 1980. (Paper presented at the Fifth International Symposium on Multivariate Analysis, University of Pittsburgh, Pittsburgh PA, 1978.) (c)

Samejima, F. Research on the multiple-choice test item in Japan: Toward the validation of mathematical models (ONRT-M3). Tokyo: Department of the Navy, Office of Naval Research, April 1980. (d)

Samejima, F. Estimation of the operating characteristics when the test informa-

tion of the old test is not constant II: Simple sum procedure of the conditional p.d.f. approach/normal approach method using three subtests of the old test (Research Report 81-2). Knoxville: University of Tennessee, Department of Psychology, July 1981.

Samejima, F., & Changas, P. S. How small the number of test items can be for the basis of estimating the operating characteristics of the discrete responses to unknown test items (Research Report 81-3). Knoxville: University of Tennessee, Department of Psychology, November 1981.

Samejima, F., & Livingston, P. Method of moments as the least squares solution for fitting a polynomial (Research Report 79-2). Knoxville: University of Tennessee, Department of Psychology, June 1979.

Shiba, S. Construction of a scale for acquisition of word meanings. Bulletin of the Faculty of Education, University of Tokyo, 1978, 17, 47-58. (In Japanese with English abstract)

Shiba, S., Noguchi, H., Haebara, T. A stratified adaptive test of verbal ability. Japanese Journal of Educational Psychology, 1978, 26, 11-20. (In Japanese with English abstract)

Tucker, L. R. Academic ability test (Research Memorandum 51-17). Princeton NJ: Educational Testing Service, 1951.